A multiple crop model ensemble for improving broad-scale yield prediction using Bayesian model averaging

Xiao Huang, Guorui Huang, Chaoqing Yu*, Shaoqiang Ni, Le Yu

Department of Earth System Science, Ministry of Education Key Laboratory for Earth System Modeling, Tsinghua University, Beijing, China

ABSTRACT

Process-based crop models are popular tools to evaluate the impact of climate change and agricultural management on crop growth. Accurate simulation of crop production over large geographic regions using an individual crop model remains challenging due to different sources of uncertainty. We present a Bayesian model averaging (BMA) method for a multiple crop-growth model ensemble to provide more reliable predictions of maize yields in Liaoning Province, northeastern China, which covers an area of 148,000 km^2 and has 2200,000 ha of maize. We apply the photosynthesis-oriented WOFOST (WOrld FOod STudy) model, the water-oriented AquaCrop model and the nitrogen-oriented DNDC (DeNitrification and DeComposition) model to independently generate original predictions of county-level maize yields. The integrated prediction is achieved using a linear combination of the three ensemble members using BMA weights. This integrated approach results in more accurate and precise predictions than any individual model over the entire province. This is because the BMA framework effectively compensates for the uncertainty of individual model simulation and takes advantage of each competing model for reliable prediction. Furthermore, the interpretation of the BMA weight values is also strengthened by comparison with regional precipitation, fertilization and radiation data. We find these values adequately fit the regional limiting factors, e.g., the AquaCrop model generally has a high weight value in counties with frequent droughts, while WOFOST is the dominant member in areas with radiation deficit. Compared with the simple average method and median estimate, the results show that the BMA framework is powerful in computing the ensemble weights and interpreting the mechanism beyond the observed data.

1. Introduction

Crop growth and food production are sensitive to climate change (Adams et al., 1998; Xiao et al., 2016). Climatic variability is expected to increase production risk (Rosenzweig et al., 2014; Fires et al., 2016), while the food demand from population growth raises concern about food security (Bloom, 2011). More accurate prediction of food production at a large scale is needed for both scientific researchers and policy makers (Vermeulen et al., 2013; Li et al., 2016). It is therefore of great importance to understand how crop growth responds to different environmental components, including climate, soil and management practices. Given the complex processes and the uncertainty considered, process-based crop models can be applied as a valuable tool to address these problems for robust decision making (Tubiello and Ewert, 2002; Rosenzweig et al., 2013).

Various process-based crop models have been developed and applied to simulate and evaluate the impacts of environmental factors, such as water stresses, nutrient deficits, and increasing temperatures

and CO_2 concentrations, on the physiological processes of crop growth. For example, some crop models, including WOFOST (Vandiepen et al., 1989; van Ittersum et al., 2003), DSSAT (Jones et al., 2003), APSIM (McCown et al., 1996), AquaCrop (Steduto et al., 2009; Raes et al., 2009a) and EPIC (Jones et al., 1991), simulate crop growth and yield development processes using carbon and water as the main driving forces. Another class of models simulates biogeochemical processes with coupled water, carbon, nitrogen and phosphorus cycles, such as Century (Gilmanov et al., 1997) and DNDC (Li et al., 1994). These biogeochemical models have been widely used for the integrated assessment of sustainable management in agro-ecosystems (Cui et al., 2014; Zhang et al., 2015). Both of these groups of models have been applied and evaluated at a field scale and across large geographic regions (de Wit and van Diepen, 2007; Hsiao et al., 2009; Li et al., 2012; Yu et al., 2014; Chen et al., 2015).

All these models are viewed as imperfect representations of real natural processes. They vary both in model structure and parameterization scheme, which are considered as the dominating

contributors to the uncertainty in crop growth prediction (Murphy et al., 2004; Duan et al., 2007). Any particular model may have strength in simulating some aspects of the crop growth process, but there is no model that is superior to others under all environmental conditions (Todorovic et al., 2009; Bassu et al., 2014). Currently, many methods are used to improve the predictive skill of crop models, including stochastic parameter optimization (Iizumi et al., 2009; Dumont et al., 2014) and ensemble driving forces (Tao and Zhang, 2013). Despite the progress made, model structure uncertainty remains an inevitable issue for individual model simulation. As a result, for large-scale prediction in a complex environment, model choice remains a major challenge for reliable prediction. Multi-model ensemble methods have become a widely accepted approach to improve prediction by adjusting various biases and taking advantage of complementary individual models (Raftery et al., 2005; Gupta et al., 2012). While multi-model ensembles have gained popularity in the fields of hydrology (Mehrotra et al., 2014; Najafi and Moradkhani, 2015), climate (Sahai et al., 2008), and economy (Kar et al., 2006; Belyaeva et al., 2014), intensive applications in crop modeling have not occurred until recently. The Agricultural Model Intercomparison and Improvement Project (AgMIP, www.agmip.org) is a good example that links the international climate, crop, and economic modeling communities to improve crop modeling performance and reduce uncertainty (Rosenzweig et al., 2013). Martre (Martre et al., 2015) used both the mean and median of 27 wheat crop model simulations in a multi-model ensemble and showed that the multi-model prediction was more accurate than the individual model results.

A challenge existing in multi-model ensemble approaches is how to weigh different models with consideration of individual model strength and variations in the crop-growth environment. Several methods have been developed to determine the weight value of each competing model. The simplest method is the simple model average (SMA) approach, which is to assign an equivalent weight value for all models, so the ensemble estimator is the mean of the model simulations. However, the weights determined by SMA have no connection with the individual model performance (Duan et al., 2007). For crop yield simulation over a large geographic region, these weights are generally inadequate for model ranking and selection and are unable to provide effective information to reveal the key limiting process or driving force of crop production under spatial variability. The median estimate (ME) is similar to SMA and uses the median for ensemble members as its prediction (Martre et al., 2015). Another method is to allocate weights to each competing model according to its fitness, and a consensus prediction can be obtained through the linear combination of individual model simulation with its corresponding weight (Thompson, 1977; Smith et al., 2004). This approach focuses on capturing the uncertainty associated with model structure, boundary conditions and physical processes, and the multi-model ensemble average has convincingly demonstrated improved prediction and reliability compared with individual model simulation. The algorithms for this approach include the artificial neural network (ANN) estimate (Shamseldin and O'Connor, 1999) and the Bayesian model averaging (BMA) method (Madigan et al., 1996; Raftery et al., 1997).

The BMA framework is especially good at assigning weights according to individual model strengths. The model weights are obtained based on Bayes' theorem by updating the prior belief with the likelihood of model prediction given the observation (Raftery et al., 1997). The posterior probabilities resulting from this rigorous statistical framework, as model weights, objectively demonstrate the difference in the reproducing ability of each model relative to observed data (Madigan et al., 1996; Hoeting et al., 1999). For crop growth simulation, a particular individual model that captures the key limiting process is more likely to fit the dynamics of a target variable and thus exhibit a higher weight value. Moreover, BMA provides additional uncertainty descriptions, including inner model variance and between-model variance, and results in more accurate and reliable predictions

than many other multi-model methods (Raftery et al., 2005). The BMA scheme has been applied in different applications including hydrological modeling (Ajami et al., 2007; Duan et al., 2007), climate projection (Miao et al., 2014), ecology (van Oijen et al., 2013) and soil-plant simulation (Wöhling et al., 2015). However, the application of BMA for yield prediction has rarely been reported. Moreover, as model ranking differs significantly due to the specific data set used in the calibration, whether the weights are representative in terms of the model predictive skills is the focus of research for BMA application (Schöniger et al., 2015; Wöhling et al., 2015).

In this study, the application of BMA for maize yield prediction was explored in Liaoning Province, China. The area of this province covers approximately 148,000 km^2, with considerable variability in climate. A total of 2200,000 ha of maize was planted in 2013. Three different crop models were applied to simulate maize growth and yield formation in this province: the photosynthesis-oriented WOFOST model, the water-oriented AquaCrop model and the nitrogen-oriented DNDC model. We aim to investigate: (i) whether the BMA scheme can improve the accuracy and reliability of yield prediction and (ii) whether the BMA weights are representative enough in explaining the key limiting factor for crop production as different models have different structures and primary driving elements.

Section 2 introduces the methodology of the BMA framework. Section 3 briefly describes the strategies for yield simulation of three models, and Section 4 introduces the study region and data required for the simulation. In Section 5, results from the three individual models and the BMA approach are compared for validation. Finally, the discussion and conclusion are presented in Section 6.

2. Bayesian model averaging (BMA)

In the BMA approach as expressed in Eq. (1), yi presents the ensemble member prediction of the specific variable of interest from the ith model, and Y is the quantity of interest. For each ensemble member, there is a conditional probability density function (PDF), pi(Y|yi), which represents the conditional PDF of Y given that the ith model in the ensemble is under consideration. Then, based on the law of total probability, the conditional PDF of the BMA prediction on all model ensemble members can be expressed as (Madigan et al., 1996; Raftery et al., 1997):

$$p(Y|y_1,...,y_n) = \sum_{i=1}^{n} w_i p_i(Y|y_i) \tag{1}$$

where n denotes the total number of competing models, and wi is the posterior probability of the ith model prediction yi, which indicates the likelihood of yi as the best forecast given the observed data. This weight shows the relative contribution of the corresponding model to the overall prediction, and we therefore obtain $\sum_{i}^{n} w_i = 1$.

The conditional PDF of pi(Y|yi) of the ensemble members is assumed to have a Gaussian distribution with a mean of aiyi + bi and a standard deviation σi (Raftery et al., 2005):

$$Y|y_i \sim N(a_i y_i + b_i, \sigma_i^2) \tag{2}$$

where ai and bi are the parameters of the linear bias-correction of the linear regression of Y on yi. The posterior BMA ensemble mean and variance can be expressed as (Raftery et al., 2005):

$$E[Y|y_1,...,y_n] = \sum_{i=1}^{n} w_i(a_i y_i + b_i) \tag{3}$$

$$Var[Y|y_1,...,y_n] = \sum_{i=1}^{n} w_i \left[(a_i y_i + b_i) - \sum_{i=1}^{n} w_i(a_i y_i + b_i) \right]^2 + \sum_{i=1}^{n} w_i \sigma_i^2 \tag{4}$$

Consequently, the BMA ensemble mean is the linear combination of the individual ensemble predictions and their weights. BMA ensembles